

Response to Comments of Referee #2

This study compares the performance of 13 GHMs and LSMs in capturing amplitude and phase of TWSC in global major rivers against GRACE data, including comparisons across climate zones and model version (R1 and R2). This detailed comparison facilitates improved parameterization model process. However, limitations of this study include overly detailed descriptions of the comparisons of different basins so that it is difficult for the reader to get to the point, and the figure lacks summarization.

Major comments:

1. The configuration of the modules of each model should be clearly stated, and some differences may be due to missing modules, e.g., snow, permafrost, groundwater, etc., which would also be useful for analyzing the causes of deviations. And, if key modules are missing does it still make sense to compare changes in TWSC between the model (GHMs and LSMs) and GRACE.

Response: Thank you for your valuable feedback on our manuscript. We appreciate your thoughtful comments and suggestions. In response to your specific points:

We agree that providing a clear and comprehensive description of the configuration of the modules in each model is essential. In our revised manuscript, we included details of key modules within each model in Table 1, including any specific modules for snow, groundwater, and other relevant components. This will help readers understand the differences and similarities between the models and their potential impact on simulation results. We acknowledge the importance of considering the potential impact of missing key modules in the GHM and LSM models. While some differences between simulations of the models and GRACE may indeed be due to these missing modules, we believe that the comparison still holds value. Because, we aim to assess the agreement and discrepancies between the models and GRACE in terms of Total Water Storage Changes (TWSC) for better understanding the limitations of this approach. By highlighting the missing modules in section 4.1, we provided insights into the potential sources of deviations and uncertainties in TWSC estimates.

2. This study did not analyze in depth the causes of amplitude and phase differences, especially 4.1 section.

Response: Thank you for your valuable feedback. We appreciate your suggestion to delve deeper into the causes of amplitude and phase differences in Section 4.1. In response to this comment, we expanded the discussion in Section 4.1 and provided a more comprehensive analysis of the factors contributing to the observed amplitude and phase differences.

3. Line67, “due to human intervention and climate change respectively”, the underestimation is due to anthropogenic interventions and climate change,

doesn't that have anything to do with model performance, shouldn't model performance be the main reason?

Response: Thank you for your comment. We rephrase the lines as “Compared to GRACE-derived TWS trends, Scanlon et al. (2018) revealed that the TWS trends of GHMs were either underestimated or had the opposite sign over numerous basins across the globe”

4. Line 89, amplitude, and phase of “polar” zone was not analyzed in result section.

Response: Thank you for your valuable comment. In this study, we primarily concentrated on analyzing the boreal, temperate, arid, and tropical zones, we did not include the polar zone in our analysis. However, we believe that exploring the amplitude and phase of the polar zone could indeed be a valuable avenue for future research to provide a more comprehensive understanding of the subject matter. We will duly consider this suggestion for future studies in this field.

5. Line 139, Why not CSR and JPL on average?

Response: We appreciate your comments and the opportunity to clarify our choice of using GRACE data from two data processing centers rather than utilizing the average. It is common in the field of Earth sciences to use data from multiple sources, and it is often encouraged to include data from different processing centers to account for potential biases and uncertainties in the measurements. Our decision to use GRACE data from two processing centers was made to enhance the robustness and reliability of our findings and better capture regional variations. We believe this approach aligns with best practices in the field and contributes to the scientific rigor of our study.

6. The difference in the length of the text in parts 3.1 and 3.2 is too large. 3.1 section over-emphasis on basin comparisons.

Response: Thank you for the reviewer's comment regarding the difference in the length of text between sections 3.1 and 3.2. We appreciate your feedback, and we will work to ensure a more balanced and consistent presentation of information in these sections.

To address this concern, we will review and revise Section 3.1 to ensure that it does not over-emphasize basin comparisons and that it aligns more evenly with Section 3.2 in terms of content length. This will help maintain a better structural balance and coherence in the paper while providing equal attention to all relevant aspects of the study.

Your input is valuable, and we will improve the overall flow and readability of our manuscript.

7. The figures are not summarizing enough, too many similar comparisons, e.g., I think Figures 5-8 should be in the Appendix, and the main results should be put in the main text, e.g., the overall results for the different climatic zones in one fig.

Response: Thank you for the reviewer's comment regarding the figures in our manuscript. We understand your concern about the number of comparisons and the desire for a more concise summarization. However, we believe that Figures 5-8 are important for understanding the detailed results and patterns in different regions and should remain in the main text.

To address your suggestion for a more concise presentation of the overall results for different climatic zones, we will work on improving the clarity of the figures and their captions to ensure that readers can easily grasp the key findings. This will help strike a balance between providing detailed regional information and presenting a clear overview of the main results.

We appreciate your feedback and are committed to enhancing the presentation of our results to make them more accessible to readers while preserving the important details provided by these figures.

8. I suggest to add the spatial distribution map of biases in amplitude and phase.

Response: Thank you for your suggestion to include spatial distribution maps of biases in amplitude and phase. We understand the importance of visualizing these biases for a comprehensive understanding of the results. However, we want to clarify that such maps have already been provided by Schellekens et al. (2017), and our study relies on their analysis in this regard. Including redundant maps in our paper would indeed be repetitive and not add significant new insights to the existing literature.

We appreciate your concern, and to ensure clarity in our paper, we will explicitly reference and acknowledge the work of Schellekens et al. (2017) for the spatial distribution maps of biases in amplitude and phase between the models and GRACE data. This will help readers access the relevant information in the cited source while maintaining the focus of our study on its unique contributions and analyses.

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Spera Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G. P.: A global water resources ensemble of hydrological models: the earth2Observe Tier-1 dataset, *Earth Syst Sci Data*, 9, 389–413, <https://doi.org/10.5194/essd-9-389-2017>, 2017.

9. Figure 1-4 suggests the addition of lines for the GHM and LSM model averages, which facilitates comparison of the two types of models

Response: Thank you for your feedback regarding Figure 1-4. We appreciate your suggestion to add lines for the GHM and LSM model averages to facilitate a clearer comparison between the two types of models. We have now incorporated these lines into the figures as per your recommendation. This enhancement should provide readers with a more comprehensive view of the model comparisons and improve the overall clarity of the presentation.

Minor comments:

1. Line 4, "(e.g., the amount and" misses the corresponding right parentheses.

Response: We appreciate your suggestion and added parenthesis in the revised manuscript.

2. Line68, "Other studies focused on the seasonal cycle of TWSC to identify" to "Other studies on the seasonal cycle of TWSC focus on identifying" is more suitable? "disparities", specifically what are the disparities?

Response: Thank you for your suggestion. We have revised the sentence as follows: "Other studies on the seasonal cycle of TWSC, such as Zhang et al. (2017), have focused on identifying disparities." The term "disparities" refers to differences or variations in four global numerical model realizations that simulate the continental branch of the global water cycle and GRACE that have been investigated in previous studies.

3. Line 75, "northern basins" is vague, please specifically point

Response: Thank you for the suggestion. We have made the requested clarification in the manuscript. Line 75 now reads, "northern high-altitude basins," to provide a more specific description of the geographic region being referred to. This should help eliminate any ambiguity and ensure a clearer understanding for the readers.

4. Line 84, "replicate water storage against the latest release (RL06) of GRACE TWSC.", this sentence indicates the result? this place is to say what is to be studied

Response: We appreciate your feedback. We rephrase the sentence to "Compare high-resolution and more optimized structured R2 models against R1 models and assess their ability to simulate TWSC variability and replicate water storage against GRACE TWSC."